Recap of Part 1

Per Kragh Andersen

Section of Biostatistics, University of Copenhagen

DSBS Course
Survival Analysis in Clinical Trials
January 2018

Overview

- Definitions and examples
- Simple estimation and testing
- Regression models
- Goodness of fit for the Cox model

`www.biostat.ku.dk/~pka/SACT18-part2`

Definitions and examples

Examples

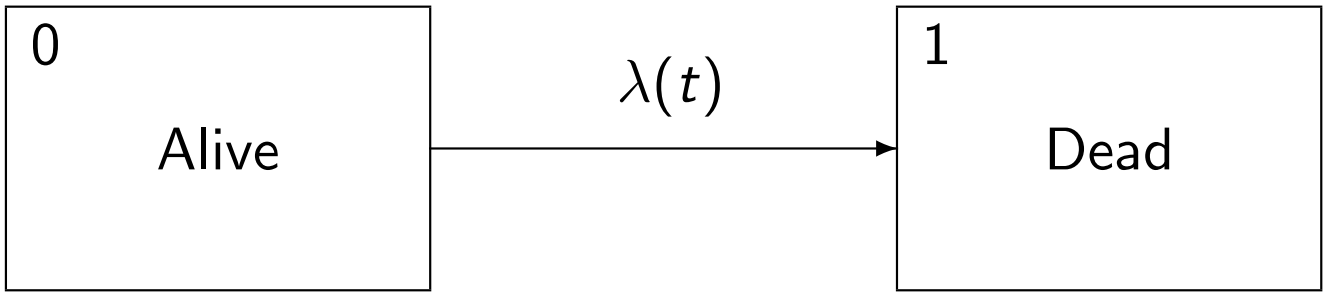
Time T to *death* or other *event* of interest from a well-defined *time origin*:

- Time from start of randomized clinical trial to death
- or ... to some *composite end-point*
- Time from randomization to occurrence of side effect
- Time from birth to death
- Time from birth to first marriage
- Time from first employment to pension
- Time from filling a cavity in a tooth to filling falls out

What is special about survival data?

- *(Right)-censoring*: For some subjects the event is not observed and we will only know an interval in which it did not occur.

The two-state model for survival data



$$\lambda(t) \approx P(\text{state 1 time } t+dt \mid \text{state 0 time } t)/dt$$

$$S(t) = P(\text{state 0 time } t) = \exp(-\int_0^t \lambda(u)du)$$

$F(t) = 1 - S(t) = P(\text{state 1 time } t)$ is the cumulative probability ('risk', cumulative incidence) of death over the interval from 0 to t

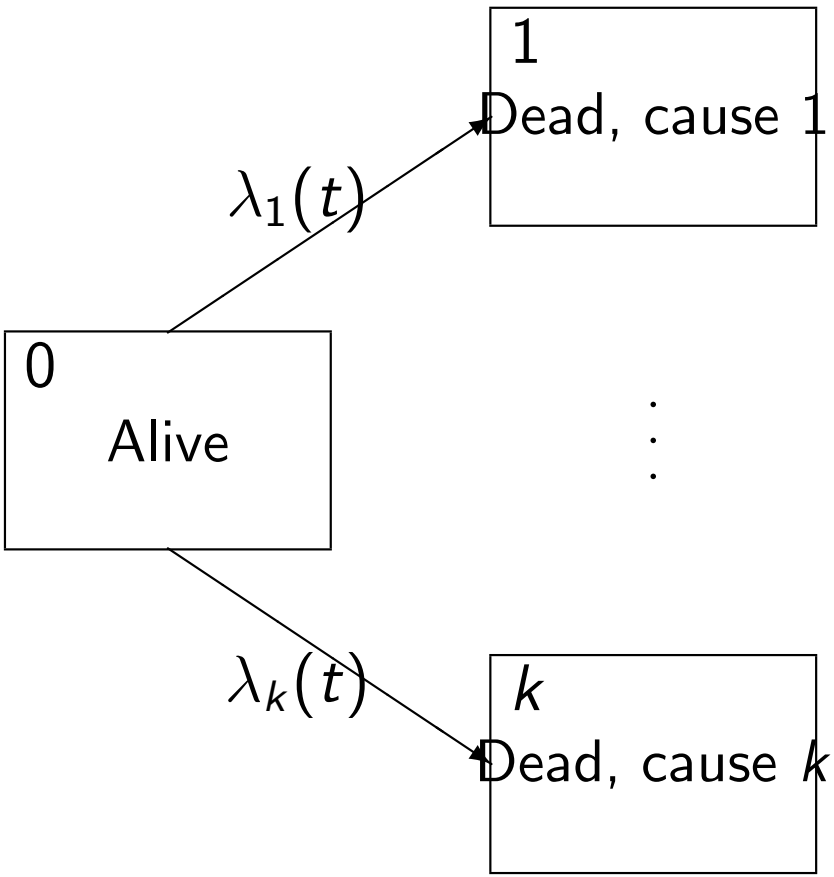
One-to-one correspondence between rate and risk

More types of event of interest

- Time from BMT to either relapse or death in remission
- Time from randomization to the occurrence of an adverse event or to withdrawal
- Time from entering a wait list until heart transplantation, or to death while on wait list, combined with time to death after transplantation
- Time from BMT to either relapse or death in remission, combined with time to death after relapse
- Time(s) from randomization to the occurrence of first, second, ... hypoglycemic event
- Time(s) from first diagnosis to first, second, ... re-admission to hospital
- Time(s) from randomization to the occurrence of first, second, ... hypoglycemic event, combined with time to an adverse event or to withdrawal
- Time(s) from first diagnosis to first, second, ... re-admission to hospital, combined with time to death

The first two are covered by the *competing risks model*.

The competing risks multi-state model



Basic parameters

Cause-specific hazards $j = 1, ..., k$ ('transition intensities'):

$$\lambda_j(t) \approx P(\text{state } j \text{ time } t + dt \mid \text{state } 0 \text{ time } t)/dt.$$

State occupation probabilities:

1. Overall survival function:

$$S(t) = P(\text{alive time } t) = \exp\Big(-\int_0^t \sum_j \lambda_j(u)du\Big).$$

2. Cumulative incidences $j = 1, ..., k$:

$$F_j(t) = P(\text{dead from cause } j \text{ before time } t) = \int_0^t S(u)\lambda_j(u)du.$$

Note that alle rates are needed to compute a single risk.

Population and sample

We are used to considering our data as a *sample* from some (target) *population*, and the parameters refer to this population.

That is no different in survival analysis, however, it is important to realize that the target population is a *complete* population, i.e., *without censoring*.

Our ambition in survival analysis is therefore to draw inference on parameters like the survival function $S(t)$, the cumulative incidence functions $F_j(t)$ or the hazard functions $\lambda_j(t)$ from a potentially completely observed population based on incomplete (censored) data.

This is quite ambitious and requires certain assumptions.

Target population; censoring

For this ambition to be feasible:

1. the complete population should be well-defined
2. censoring should not leave us with a biased sample

Requirement 1 means that, for the simple survival data situation, the event under study should happen for every one in the population. In the competing risks model *one* of the events under study should happen for every one in the population

Thus, we need to distinguish between situations where there are *no competing risks* and where there *are* competing risks, and one should note that the 'box and arrows' diagrams illustrate the *complete* population.

Independent censoring

Requirement 2 is the assumption of *independent censoring* (by some denoted *non-informative* censoring).

This means that individuals censored at any given time t should not be a biased sample of those who are *at risk* at time t.

Stated in other words: the hazard function $\lambda_j(t)$ gives the type j event rate at time t, i.e. the failure rate given that the subject is still alive ($T > t$).

Independent censoring then means that the extra information that the subject is not only alive, but also uncensored at time t does not change the failure rate.

Independent censoring cannot be tested from the available data.

The PBC-3 trial in liver cirrhosis

Lombard et al. (1993, *Gastroenterology*)

- Multi-centre randomized trial in patients with primary biliary cirrhosis.
- Patients ($n = 349$) recruited 1 Jan, 1983 - 1 Jan, 1987 from six European hospitals and randomized to CyA (176) or placebo (173).
- Followed until death or liver transplantation (no longer than 31 Dec, 1989); CyA: 30 died, 14 were transplanted; placebo: 31 died, 15 were transplanted; 4 patients were lost to follow-up before 1989.
- Primary outcome variable: time to death, incompletely observed due to: liver transplantation, loss to follow-up, alive 31 Dec, 1989
- In some analyses, the outcome is defined as 'time to failure of medical treatment', i.e. to the composite end-point of either death or liver transplantation

EBMT example

- Data from the European group for Blood and Marrow Transplantation (EBMT)
- All (3982) chronic myeloid leukemia (CML) patients with an allogeneic stem cell transplantation from an HLA-identical sibling or a matched unrelated donor during the years 1997–2000.
- Study effect of EBMT risk score with values 0–7, here grouped into five groups: 0, 1 ($n = 506$), 2 ($n = 1159$), 3 ($n = 1218$), 4 ($n = 745$), and 5, 6, 7 ($n = 354$).
- Points obtained from: donor type (2), stage (3), age (3: 20,40), female-to-male (2), time from diagnosis (2: 12 mo.)
- Failure from transplantation may either be due to relapse or to non-relapse mortality (NRM). Often these two endpoints are taken together to relapse-free survival (RFS), which is the time from transplantation to either relapse or death, whichever comes first.

Example: Drug discontinuations as competing risks

Data on time to drug discontinuation for different reasons in a 1-year RCT (n=559) of active drug vs placebo.

- `week` Time in study
- `state`
  - 0 = Complete on drug (406)
  - 1 = Adverse event (42)
  - 2 = Withdrew consent (62)
  - 3 = Other (49)
- `drug`
  - 0 = Placebo (188)
  - 1 = Active (371)

Malignant melanoma

- 205 patients with malignant melanoma (skin cancer) were operated at Odense University Hospital between 1962 and 1977
- All patients had radical operation, i.e. no treatment variable relevant here. Purpose: study prognostic factors like sex, age, thickness of tumor, ulceration
- By the end of 1977: 57 had died from the disease, 14 had died from other causes, and 134 were still alive
- Primary outcome variable: survival time from operation, but also mortality from disease is of interest
- Primary outcome variable incompletely observed due to end of follow up (and death from other causes)

Simple estimation and testing

The Kaplan-Meier estimator

Both for simple survival data and for competing risks we let $0 < t_1 < t_2 < \ldots$ be the distinct failure or censoring times, $d_j(t_1), d_j(t_2), \ldots$ the total number of failures of type j observed at those times (typically 0 or 1), and $Y(t_1), Y(t_2), \ldots$ the number of subjects *at risk* at (i.e., just before) those times.

Then the Kaplan-Meier estimator (for $t_i \leq t < t_{i+1}$) is (where $d(t) = \sum_j d_j(t)$):

$$\begin{aligned} \widehat{S}(t) &= (1 - \frac{d(t_1)}{Y(t_1)})(1 - \frac{d(t_2)}{Y(t_2)}) \cdots (1 - \frac{d(t_i)}{Y(t_i)}) \\ &= \prod_{t_i \leq t} (1 - \frac{d(t_i)}{Y(t_i)}). \end{aligned}$$

The Nelson-Aalen estimator

This estimator of the integrated hazard function $\Lambda_j(t)$ builds on the same idea as the Kaplan-Meier estimator: estimate

$$\lambda_j(t)dt \approx P(T \leq t + dt, D = j \mid T > t) \text{ by } \frac{d_j(t_i)}{Y(t_i)} \text{ when } t = t_i$$

(where D is the failure indicator). That is,

$$\widehat{\Lambda}_j(t) = \sum_{t_i \leq t} \frac{d_j(t_i)}{Y(t_i)}.$$

Note how censored observations are used for both K-M and N-Aa: a subject censored at t_i gives rise to *no jump* in the estimator but contributes to the size, $Y(t)$ of the risk set for $t \leq t_i$.

For simple survival data, K-M is the *product-integral* of N-Aa.

The Aalen-Johansen estimator

For simple survival data, the *cumulative incidence* $F(t) = 1 - S(t)$ is estimated by '1-K-M'.
With competing risks, the Aalen-Johansen estimator for

$$F_j(t) = \int_0^t S(u)\lambda_j(u)du$$

builds on *plug-in*:

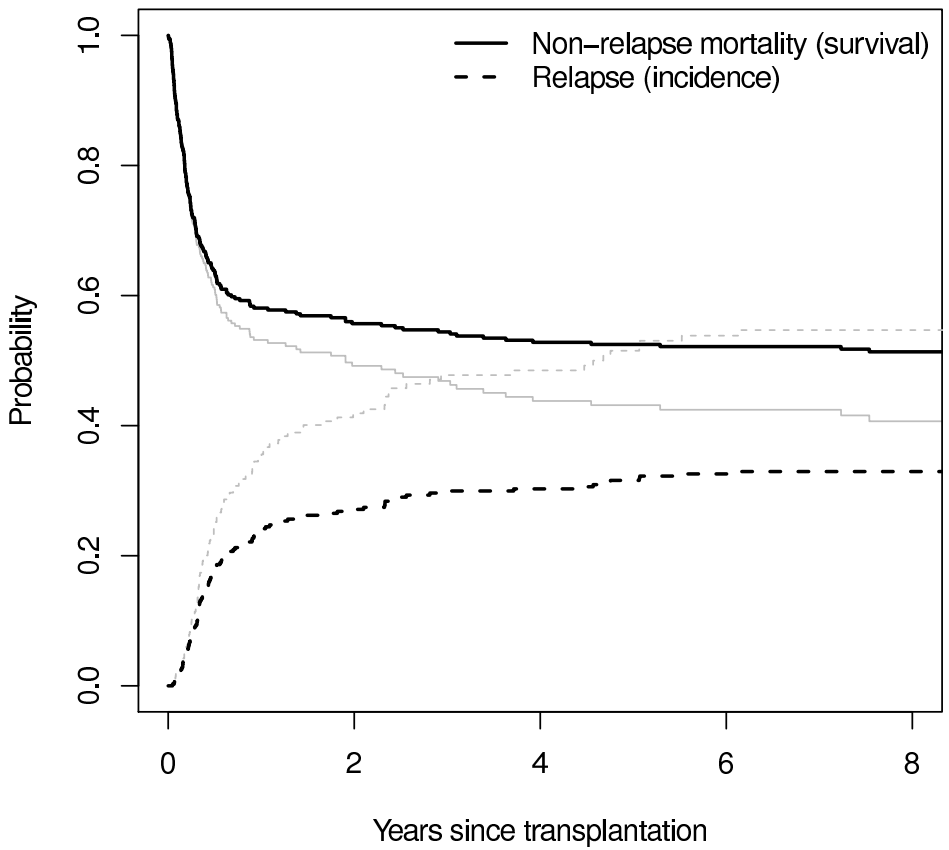
$$\widehat{F}_j(t) = \sum_{t_i \le t} \widehat{S}(t_{i-1}) \frac{d_j(t_i)}{Y(t_i)}.$$

Recall that the '1-K-M estimator based only on cause j events':

$$1 - \widehat{S}_j(t) = 1 - \prod_{t_i \le t} (1 - \frac{d_j(t_i)}{Y(t_i)})$$

is *upward biased*: $F_j(t) \le 1 - S_j(t)$.

Cumulative incidence curves: EBMT risk group 5-7



The logrank test

We want to compare hazard functions $\lambda_1(t)$ and $\lambda_2(t)$ in two groups. (NB: '1' and '2' now refer to two *groups* and not to causes of death.) However, the tests are equally applicable for cause-specific hazards when there are competing risks.

Counting process notation: In group s we have: $N_s(t) =$ number of observed events in $[0, t]$ (of the relevant type), $Y_s(t) =$ number at risk just before time t.
Nelson-Aalen estimators for $\Lambda_s(t) = \int_0^t \lambda_s(u)du$:

$$\widehat{\Lambda}_s(t) = \int_0^t \frac{1}{Y_s(u)} dN_s(u), \quad s = 1, 2.$$

Idea in general test statistic: look at K-weighted differences between increments in Nelson-Aalen estimators:

$$U(t) = \int_0^t K(u)\big(d\widehat{\Lambda}_1(u) - d\widehat{\Lambda}_2(u)\big).$$

The logrank test

Different choices of $K(\cdot)$ provide different tests with different properties.

The most common choice is

$$K(t) = \frac{Y_1(t)Y_2(t)}{Y_1(t) + Y_2(t)}$$

leading to

$$U(t) = N_1(t) - \int_0^t \frac{Y_1(u)}{Y_1(u) + Y_2(u)}(dN_1(u) + dN_2(u)).$$

Evaluated at $t = \infty$ we get the *logrank test*:

$$U(\infty) = \text{‘Observed’ - ‘Expected’ (in group 1).}$$

Likelihood

Data: $(\widetilde{T}_i, D_i), \quad i = 1, ..., n$ where $D_i = j, j = 1, ..., k$ if observed failure from cause j, $D_i = 0$ if censored. (Special case: simple survival data, $k = 1$ cause).
The likelihood can be written as a product over causes, j:

$$\begin{aligned} L &= \prod_{i=1}^{n} S(\widetilde{T}_i) \prod_{j=1}^{k} (\lambda_j(\widetilde{T}_i))^{I(D_i=j)} \\ &= \prod_{i=1}^{n} \left(\exp(-\sum_{j=1}^{k} \Lambda_j(\widetilde{T}_i))\right) \prod_{j=1}^{k} (\lambda_j(\widetilde{T}_i))^{I(D_i=j)} \\ &= \prod_{j=1}^{k} \left(\prod_{i=1}^{n} \exp(-\Lambda_j(\widetilde{T}_i))(\lambda_j(\widetilde{T}_i))^{I(D_i=j)}\right). \end{aligned}$$

Piecewise constant hazard

The hazard function is $\lambda_j(t) = \lambda_{j\ell}$ when $s_{\ell-1} \leq t < s_\ell$ for pre-specified intervals, $0 = s_0 < s_1 < \ldots < s_L = \infty$.
The maximum likelihood estimator is most easily expressed in counting process notation:

$$N_j(t) = \sum_i I(\widetilde{T}_i \leq t, D_i = j), \quad Y(t) = \sum_i I(\widetilde{T}_i \geq t).$$

Then

$$\widehat{\lambda}_{j\ell} = \frac{N_j(s_\ell) - N_j(s_{\ell-1})}{\int_{s_{\ell-1}}^{s_\ell} Y(t)dt},$$

i.e., number of cause j failures in interval ℓ divided by the total time at risk in interval ℓ.

Regression models

The Cox regression model

In the Cox model, the cause j hazard for individual $i = 1, \ldots, n$, is

$$\begin{aligned} \lambda_{ji}(t) &= \lambda_{j0}(t) \exp(\beta_{j1} X_{i1} + \cdots + \beta_{jp} X_{ip}) \\ &= \lambda_{j0}(t) \exp(\beta_j^{\mathsf{T}} X_i) \end{aligned}$$

where $\beta_{j1}, \ldots, \beta_{jp}$ are regression parameters and $X_{i1}, ..., X_{ip}$ are the covariate values for individual i. It is seen that hazards are assumed to be *proportional*.

$\lambda_{j0}(t)$ is the cause j *baseline hazard*, and no assumptions are made about its shape.

Time t is the chosen time-variable, e.g. time since randomization or age or disease duration.

Cox's partial likelihood function

Cox's partial likelihood function is

$$L(\beta) = \prod_j \prod_{i=1}^{n} \left(\frac{\exp(\beta_j^{\mathsf{T}} X_i)}{\sum_{\ell \in R(\widetilde{T}_i)} \exp(\beta_j^{\mathsf{T}} X_\ell)} \right)^{I(D_i = j)},$$

where $R(t)$ is the *risk set* at time t.

The partial likelihood function may be obtained from the general likelihood (for competing risks) by profiling out the baseline hazard function(s) $\lambda_{j0}(t)$.

Estimates of the parameters are obtained by maximizing $L(\beta)$ and the usual type of large-sample likelihood properties apply.

The score function

The log-likelihood is

$$\ell(\beta) = \log(L(\beta)) = \sum_j \sum_{i=1}^{n} I(D_i = j)\big(\beta_j^{\mathsf{T}} X_i - \log \sum_{\ell \in R(\widetilde{T}_i)} \exp(\beta_j^{\mathsf{T}} X_\ell)\big)$$

and the *score* is

$$U(\beta) = \frac{d}{d\beta}\ell(\beta) = \sum_j \sum_{i=1}^{n} I(D_i = j)\big(X_i - \frac{\sum_{\ell \in R(\widetilde{T}_i)} X_\ell \exp(\beta_j^{\mathsf{T}} X_\ell)}{\sum_{\ell \in R(\widetilde{T}_i)} \exp(\beta_j^{\mathsf{T}} X_\ell)}\big)$$

$$= \sum_j \sum_i \int_0^\infty \big(X_i - \frac{\sum_\ell Y_\ell(t) X_\ell \exp(\beta_j^{\mathsf{T}} X_\ell)}{\sum_\ell Y_\ell(t) \exp(\beta_j^{\mathsf{T}} X_\ell)}\big)\, dN_{ji}(t).$$

$$= \sum_j \sum_i \int_0^\infty \big(X_i - \bar{X}(t)\big)\, dN_{ji}(t).$$

The Breslow estimator

The cumulative cause j baseline hazard $\Lambda_{j0}(t) = \int_0^t \lambda_{j0}(s)ds$ can be estimated by the Breslow estimator

$$\widehat{\Lambda}_{j0}(t) = \sum_{\widetilde{T}_i \leq t} \frac{I(D_i = j)}{\sum_{\ell \in R(\widetilde{T}_i)} \exp(\widehat{\beta}_j^{\mathsf{T}} X_\ell)},$$

$$= \int_0^t \frac{dN_{ji}(u)}{\sum_\ell Y_\ell(u) \exp(\widehat{\beta}_j^{\mathsf{T}} X_\ell)},$$

where $\widehat{\beta}_j$ is the maximum likelihood estimate of β_j. With no covariates, the Breslow estimator is the Nelson-Aalen estimator

$$\widehat{\Lambda}_{j0}(t) = \sum_{\widetilde{T}_i \leq t} \frac{I(D_i = j)}{Y(\widetilde{T}_i)}.$$

The stratified Cox model

The proportional hazards assumption for a categorical covariate may be relaxed by considering the *stratified Cox model*.
For survival data (i.e., for simplicity we consider the situation without competing risks) the model is:

$$\lambda_i(t) = \lambda_{s0}(t)\exp(\beta^{\mathsf{T}} X_i), \text{ for } i \text{ in stratum } s, \quad s = 1, ..., k.$$

In the simplest stratified model, the regresison parameter β is the same for all s (an assumption which may be relaxed). If all the βs are stratum-specific then the model is the same as what we get by fitting a Cox model separately for each s '`PHREG BY` s'.

The Cox partial likelihood is a product over strata and the cumulative baseline hazard $\Lambda_{s0}(t)$ may be estimated by a Breslow estimator.

Predicted probabilities from Cox model for survival data

We have the relationship between survival and hazard functions

$$S(t \mid X) = [\exp(-\Lambda_0(t))]^{\exp(\beta^{\mathsf{T}} X)} = [S_0(t)]^{\exp(\beta^{\mathsf{T}} X)}.$$

The predicted survival probabilities from a Cox model for covariates X are

$$\widehat{S}(t, X) \quad = \quad \widehat{S}_0(t)^{\exp(\hat{\beta}^{\mathsf{T}} X)},$$

where

$$\widehat{S}_0(t) = \exp(-\widehat{\Lambda}_0(t)),$$

using the Breslow estimator for $\Lambda_0(t)$ (or alternative estimators for $S_0(t)$).

Estimation of cumulative incidences from hazards

Estimate $F_j(t \mid X)$ by plug-in:

$$\widehat{F}_j(t \mid X) = \int_0^t \widehat{S}(u- \mid X) d\widehat{\Lambda}_j(u \mid X).$$

Here,

$$\widehat{\Lambda}_j(u \mid X) = \widehat{\Lambda}_{j0}(u) \exp(\widehat{\beta}_{j1} X_1 + ... + \widehat{\beta}_{jp} X_p)$$

is the cumulative cause-j-hazard estimate from the Cox model and $\widehat{S}(u \mid X)$ the Cox model based estimator for the overall survival function, e.g.,

$$\widehat{S}(u \mid X) = \exp\left(- \sum_j \widehat{\Lambda}_j(u \mid X) \right),$$

or, preferably, the corresponding product-integral estimator.

Cumulative incidences from cause-specific Cox models

Important to notice:

- The Cox models impose a simple structure between covariates and *rates*.
- Due to the non-linear relationship between rates and risks in a competing risks model, this simple relationship does not carry over to the cumulative incidences.
- In particular, the way in which a covariate affects a rate can be different from the way in which it affects the corresponding risk: this will depend on how it affects the rates for the competing causes.
- In the EBMT example we saw this phenomenon when stydying group 2 vs. 0,1, relapse

Cumulative incidence regression models

The fact that plugging-in cause-specific hazard models does not provide parameters that in a simple way describe the relationship between covariates and cumulative incidences has led to the development of direct regression models for the cumulative incidences: 'marginal' models.
The most widely used such model is the *Fine-Gray* model. Recall from a Cox model for all-cause mortality that:

$$\log(-\log(1 - F(t \mid X))) = \log(\Lambda_0(t)) + \beta^{\mathsf{T}} X.$$

Fine & Gray (1999, *JASA*) studied the similar model for a cumulative incidence:

$$\log(-\log(1 - F_j(t \mid X))) = \log(\widetilde{\Lambda}_{0j}(t)) + \widetilde{\beta}_j^{\mathsf{T}} X.$$

The Fine-Gray model

The resulting derivative $\widetilde{\lambda}_j(t)$ is denoted the *sub-distribution hazard* and the Fine-Gray model is thus a proportional sub-distribution hazards model. However, a problem is that, while the hazard function has the useful "rate" interpretation:

$$\lambda(t) \approx P(\text{death before } t + dt \mid \text{ alive } t)/dt, \quad dt \text{ small},$$

and so has the cause-specific hazard:

$$\lambda_1(t) \approx P(\text{death from cause 1 before } t+dt \mid \text{ alive } t)/dt, \quad dt \text{ small},$$

the sub-distribution hazard has *not*. Thus

$$\widetilde{\lambda}_1(t) \approx P(\text{death from cause 1 before } t + dt \mid$$
either alive at t *or* death from a competing cause by $t)/dt, \quad dt$ small.

Estimation with complete data

With no censoring, Fine and Gray defined the cause j "risk set"

$$\widetilde{R}_j(t) = \{i : (T_i \geq t) \text{ or } (T_i \leq t, D_i \neq j)\}$$

and $\widetilde{\beta}_j$ is estimated by the partial likelihood score equation

$$U_j(\widetilde{\beta}_j) = \sum_i I(D_i = j)\Big(X_i - \frac{\sum_{\ell \in \widetilde{R}_j(T_i)} X_\ell \exp(\widetilde{\beta}_j^{\mathsf{T}} X_\ell)}{\sum_{\ell \in \widetilde{R}_j(T_i)} \exp(\widetilde{\beta}_j^{\mathsf{T}} X_\ell)}\Big) = 0$$

corresponding to replacing times of failure from causes other than j by $+\infty$.

Using counting process notation

$$U_j(\widetilde{\beta}_j) = \sum_i \int_0^\infty \Big(X_i - \frac{\sum_\ell \widetilde{Y}_{j\ell}(t) X_\ell \exp(\widetilde{\beta}_j^{\mathsf{T}} X_\ell)}{\sum_\ell \widetilde{Y}_{j\ell}(t) \exp(\widetilde{\beta}_j^{\mathsf{T}} X_\ell)}\Big) dN_{ji}(t) = 0$$

with $\widetilde{Y}_{ji}(t) = 1 - N_{ji}(t-)$, i.e. no cause j failure by time t.

Estimation with censored data

With known (e.g., "administrative") censoring (at U_i), the cause j risk set is replaced by

$$\widetilde{R}_j(t) = \{i : (T_i \wedge U_i \geq t) \text{ or } (T_i \leq t, D_i \neq j, U_i \geq t)\},$$

i.e. $\widetilde{Y}_{ji}(t)$ is replaced by $\widetilde{Y}_{ji}(t)I(U_i \geq t)$, that is, no cause j failure and no censoring by time t.
With general censoring, an Inverse Probability of Censoring Weighted (IPCW) score equation is used and to do this, a model for censoring is needed.
In the simplest case, one uses the 'Kaplan-Meier for censoring', that is, estimating $G(t) = P(U > t)$ (in this analysis 'failures are censorings').
If censoring depends on covariates then a model for $G(t \mid X) = P(U > t \mid X)$ is needed for the weights, e.g. a Cox model.

Estimation with censored data

For this to work, we define the weights

$$w_i(t) = I(U_i \geq T_i \wedge t)\frac{\widehat{G}(t)}{\widehat{G}(\widetilde{T}_i \wedge t)},$$

and Fine and Gray showed that if (in the simplest case) U_i is independent of (T_i, D_i) and X_i then the 'score' equation

$$\widetilde{U}_j(\widetilde{\beta}_j) = \sum_i \int_0^\infty (X_i - \frac{\sum_\ell w_\ell(t)\widetilde{Y}_{j\ell}(t)X_\ell \exp(\widetilde{\beta}_j^{\mathsf{T}} X_\ell)}{\sum_\ell w_\ell(t)\widetilde{Y}_{j\ell}(t)\exp(\widetilde{\beta}_j^{\mathsf{T}} X_\ell)}) w_i(t) dN_{ji}(t) = 0$$

is an *unbiased estimating equation* yielding consistent estimates of $\widetilde{\beta}$.

Estimation with censored data

The resulting weights are as follows:

$t, \widetilde{T}_i$	Status	$I(U_i \geq T_i \wedge t)$	$\widetilde{Y}_{ji}(t)$	$w_{ij}(t)$
$t \leq \widetilde{T}_i$	$D_i = 0$	1	1	1
	$D_i = j$	1	1	1
	$D_i \neq 0, j$	1	1	1
$t > \widetilde{T}_i$	$D_i = 0$	0	1	0
	$D_i = j$	1	0	$\widehat{G}(t)/\widehat{G}(\widetilde{T}_i)$
	$D_i \neq 0, j$	1	1	$\widehat{G}(t)/\widehat{G}(\widetilde{T}_i)$

After an observed time of failure (from a cause $\neq j$), a subject gets a smaller and smaller weight as time passes (and it is, therefore, less and less likely that the subject would still be uncensored).

Robust variances (in general)

Let $\widehat{\theta}$ be the solution to the unbiased estimating equation $U(\theta) = 0$.
Taylor expansion of $U(\cdot)$ around the true value, θ_0 yields:

$$U(\theta) = U(\theta_0) + U'(\theta^*)(\theta - \theta_0)$$

with θ^* on the line segment between θ and θ_0. Inserting $\widehat{\theta}$ and re-arranging we get:

$$n^{1/2}(\widehat{\theta} - \theta_0) \approx -U'(\theta_0)^{-1}(n^{-1/2}U(\theta_0)).$$

A CLT for $n^{-1/2}U(\theta_0)$ (sum of independent terms) gives a CLT for $n^{1/2}(\widehat{\theta} - \theta_0)$ and the robust ('sandwich') variance estimate is

$$(U'(\widehat{\theta})^{-1})^{\mathsf{T}}(\sum_i U_i(\widehat{\theta})^{\mathsf{T}} U_i(\widehat{\theta}))U'(\widehat{\theta})^{-1}.$$

Estimating the baseline sub-distribution hazard

There is a 'Breslow-type' estimator for $\widetilde{\Lambda}_{0j}(t)$.
With the weights $w_i(t)$ the estimator is

$$\widehat{\widetilde{\Lambda}}_{0j}(t) = \sum_i \int_0^t \frac{w_i(u) dN_{ji}(u)}{\sum_\ell w_\ell(u) \widetilde{Y}_{j\ell}(u) \exp(\widetilde{\beta}_j^\mathsf{T} X_\ell)}.$$

Fine and Gray (1999) provided asymptotics for the estimator and discussed covariate-specific predicted cumulative incidences based on the model:

$$\widehat{F}_j(t \mid X) = 1 - \exp(-\widehat{\widetilde{\Lambda}}_{0j}(t) \exp(\widetilde{\beta}_j^\mathsf{T} X)).$$

This is, in fact, the cumulative incidence estimator provided by `SAS PROC PHREG`.

Rates vs. risks?

Competing risks 'analogy':

Gladiators

Suppose that a gladiator may lose via two quite different mechanisms: lions or fellow gladiators.

When training a gladiator, he should both be prepared to face a lion or a fellow gladiator, and special skills may be needed to face a lion (even in the presence of the competing risk) and, similarly, special skills may be needed to beat a fellow gladiator. The cause-specific hazards describe how these mechanisms depend on properties and equipment of the gladiator.

For Caesar to predict the number of remaining gladiators still around at time t, and how many are lost due to lions or fellow gladiators, both risks must be considered. He needs the cumulative incidences given the distribution of properties and equipment of the population of gladiators.

Goodness of fit for the Cox model

Checking assumptions for the linear predictor

This is not different from any other model with a linear predictor (e.g., linear or logistic regression).

- No interaction between X_{i1} and X_{i2} can be tested by adding suitable interaction terms to the model
- Linearity for a quantitative X may be tested by adding, e.g., quadratic terms X^2 or linear splines to the model. For chosen cut-points, say a_1, a_2, add

$$(X - a_1)I(X > a_1) \text{ and } (X - a_2)I(X > a_2)$$

  to a model that also includes X. The dose-response relationship between X and the log(hazard) is then a broken straight line and coefficients for the linear splines give the change in slope at each cut-point.
- Martingale residuals

Checking proportional hazards

- Graphical methods based on the stratified model, e.g. by ploting $\log(\widehat{\Lambda}_{0s}(t))$ against t (or $\log(t)$) for each stratum s and see if curves have constant vertical distance
- Modeling time-dependent effects via interactions with functions of time, e.g. add $X \cdot I(t > \tau)$ or $X \cdot \log(t)$ to a model including X
- Score ('Schoenfeld') residuals

`ASSESS` statement in `PROC PHREG` uses martingale or score residuals (code below).

Example: malignant melanoma, death from disease

Cox model with `sex, thickness, ulceration`:

Covariate	$\widehat{\beta}$	SD
Sex (m vs. f)	0.459	0.267
Thickness (*mm*)	0.113	0.038
Ulceration (yes vs. no)	1.170	0.311

Linearity of tumor thickness

Add the linear splines

$$(\text{thickness} - a_j)I(\text{thickness} \geq a_j)$$

for some cutpoints $a_1, a_2, \cdots$, e.g. 2 and 5 *mm*, to the previous model.

Covariate	$\widehat{\beta}$	*SD*
Sex	0.457	0.289
Thickness	1.006	0.440
Ulceration	0.884	0.326
$(\text{thickness} - 2)I(\text{thickness} \geq 2 \text{ mm})$	-0.968	0.530
$(\text{thickness} - 5)I(\text{thickness} \geq 5 \text{ mm})$	0.042	0.205

Likelihood ratio test statistic for the reduction of the model: 5.36 (2 d.f.), $P = 0.07$.

Linearity of log(tumor thickness)

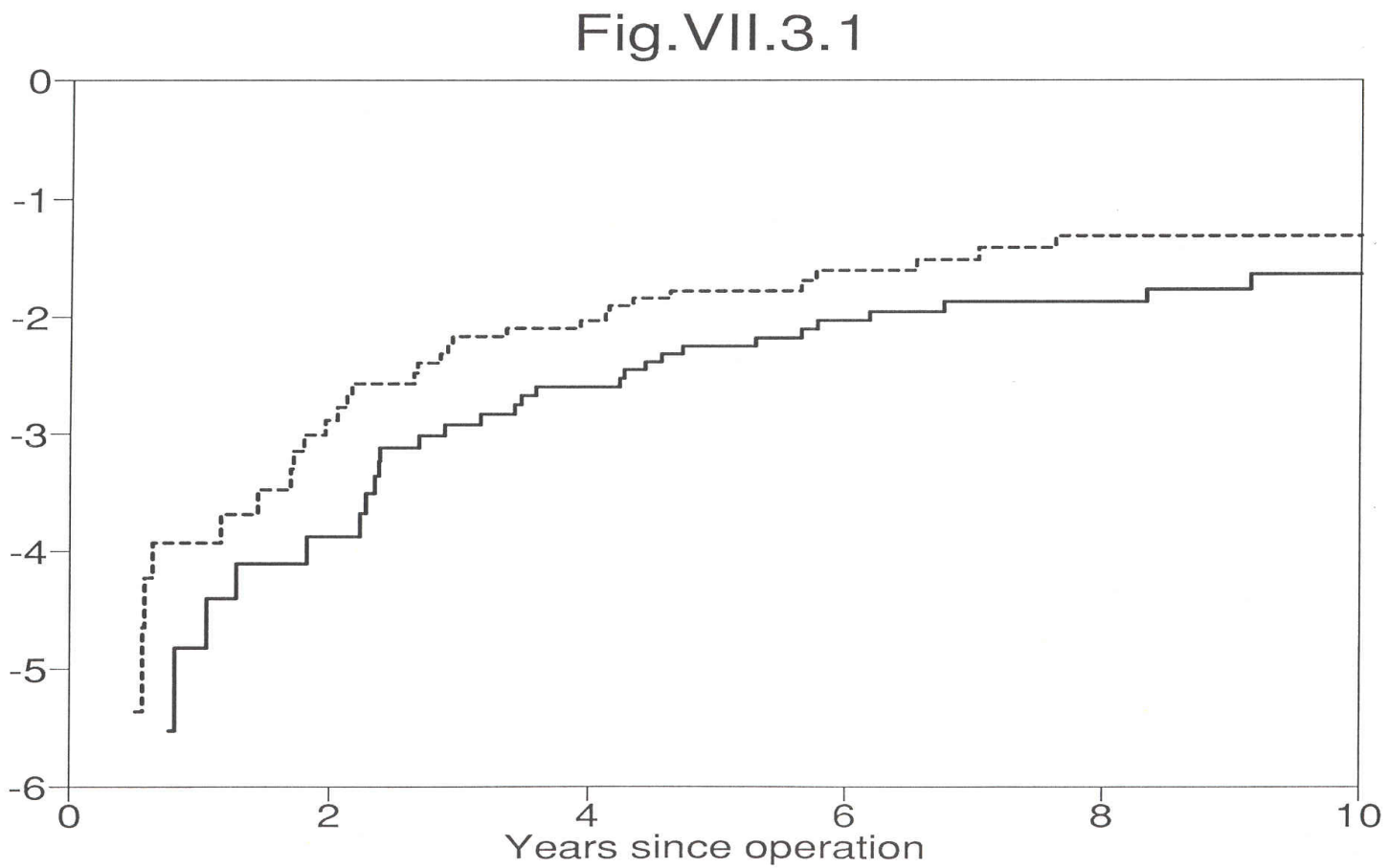
Replace thickness by log(thickness):

Covariate	$\widehat{\beta}$	SD	$\widehat{\beta}$	SD
Sex	0.401	0.285	0.381	0.271
Ulceration	0.884	0.328	0.939	0.324
log(thickness)	0.965	0.525	0.576	0.179
$(\log(\text{thickness}) - \log 2)I(\text{thickness} \geq 2)$	-0.536	0.894	—	—
$(\log(\text{thickness}) - \log 5)I(\text{thickness} \geq 5)$	-0.077	1.029	—	—

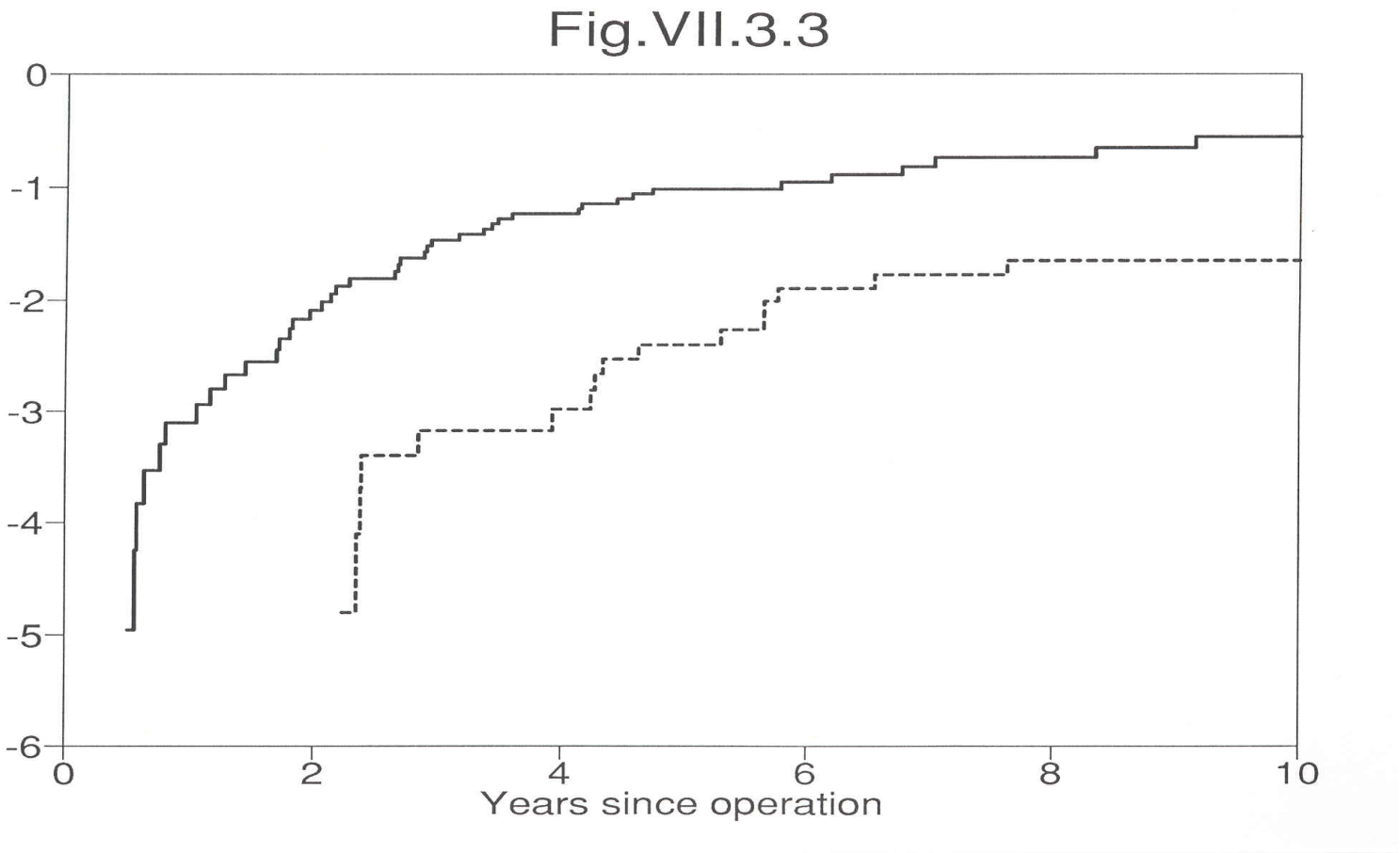
Likelihood ratio test statistic for the reduction of the model: 0.86 $P = 0.65$.

Proportional hazards: to make the plots for `thickness`, data have been stratified at 2 and 5 *mm*.

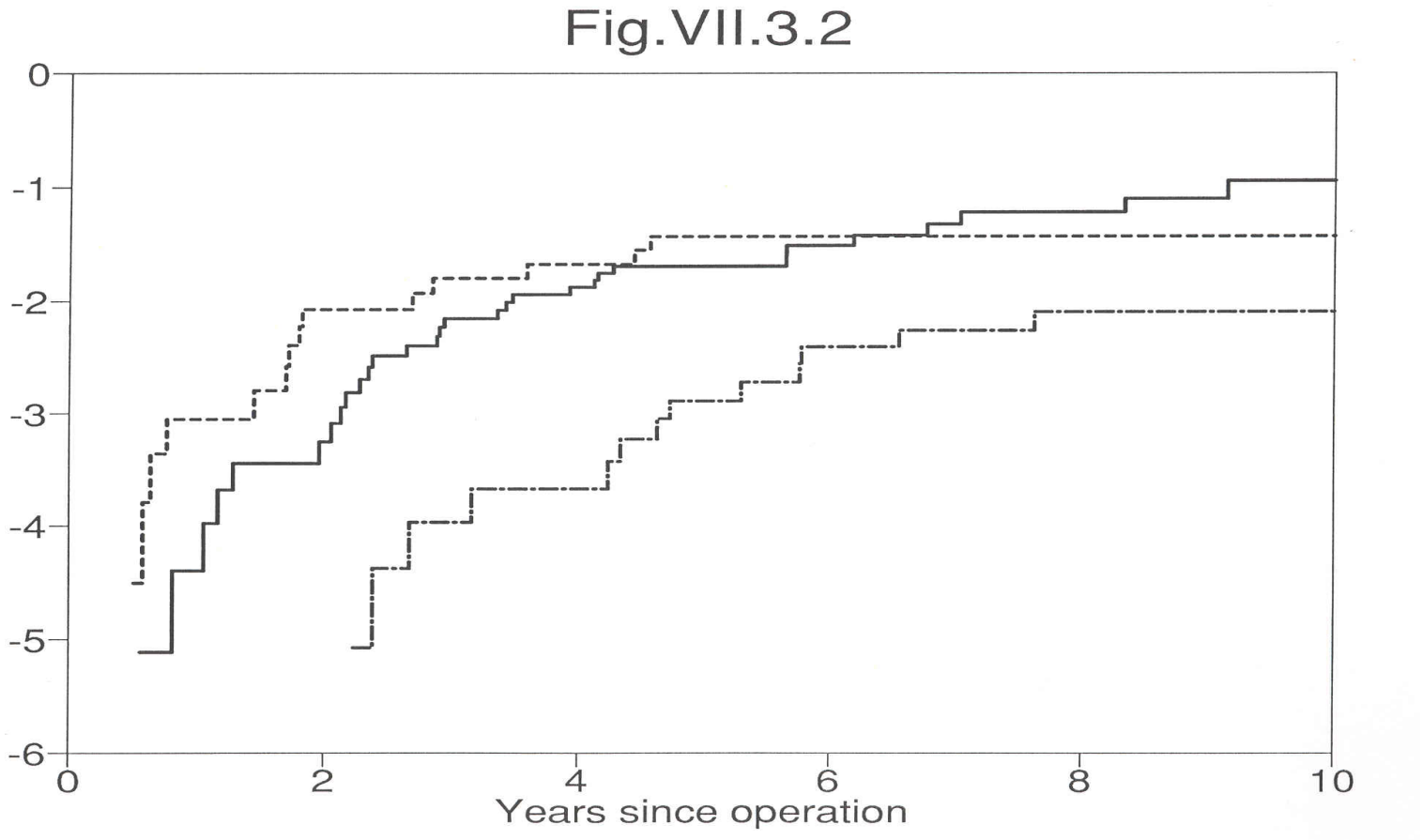
Sex



Ulceration



Tumor thickness



Example: malignant melanoma, death from disease

Add, in turn, the time-dependent covariates

1. $z_5(t) = \text{sex} \cdot (\log(t) - 7)$,
2. $z_6(t) = \text{ulceration} \cdot (\log(t) - 7)$,
3. $z_7(t) = I(2 \text{ mm} \leq \text{thickness} < 5 \text{ mm})(\log(t) - 7)$,
   $z_8(t) = I(\text{thickness} \geq 5 \text{ mm})(\log(t) - 7)$,

Time, t is in days and $7 \simeq \log(3 \cdot 365)$

Covariate	$\widehat{\beta}$	(SD)	$\widehat{\beta}$	(SD)	$\widehat{\beta}$	(SD)
Sex	0.352	(0.276)	0.372	(0.270)	0.419	(0.271)
Ulceration	0.932	(0.324)	1.048	(0.360)	0.960	(0.326)
log(thickness)	0.582	(0.180)	0.576	(0.181)	0.547	(0.191)
$z_5(t)$	-0.408	(0.394)				
$z_6(t)$			-1.189	(0.589)		
$z_7(t)$					-0.677	(0.594)
$z_8(t)$					-1.513	(0.600)

Tests for proportional hazards

1. sex: 1.12, 1 d.f., $P = 0.29$.
2. ulceration: 5.20, 1 d.f., $P = 0.02$.
3. thickness: 8.28, 2 d.f., $P = 0.02$.

For a model *stratified* by ulceration

Covariate	$\widehat{\beta}$	(SD)	$\widehat{\beta}$	(SD)	$\widehat{\beta}$	(SD)
Sex	0.402	(0.271)	0.360	(0.270)		
log(thickness)	0.538	(0.191)	0.560	(0.178)	0.589	(0.175)
$z_7(t)$	-0.367	(0.663)				
$z_8(t)$	-1.201	(0.636)				

Here, the P-value for proportional hazards for log(thickness) is 0.06 and the P-value for the effect of sex is 0.18.

Note that each test for a given variable assumes that the model fits for the other variables in the model.

Martingale residuals

Consider a given event (no cause j in notation though the event could be a given cause of failure).
Recall that the counting process for subject i is

$$N_i(t) = I(\widetilde{T}_i \leq t, D_i = 1)$$

and counts $+1$ at the observed time of failure. Note that $N_i(\infty) = D_i$. Let $Y_i(t) = I(\widetilde{T}_i \geq t)$ be the *at-risk indicator* for subject i. Then the *martingale residual* (process) is

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) \exp(\beta^{\mathsf{T}} X_i) \lambda_0(u) du$$

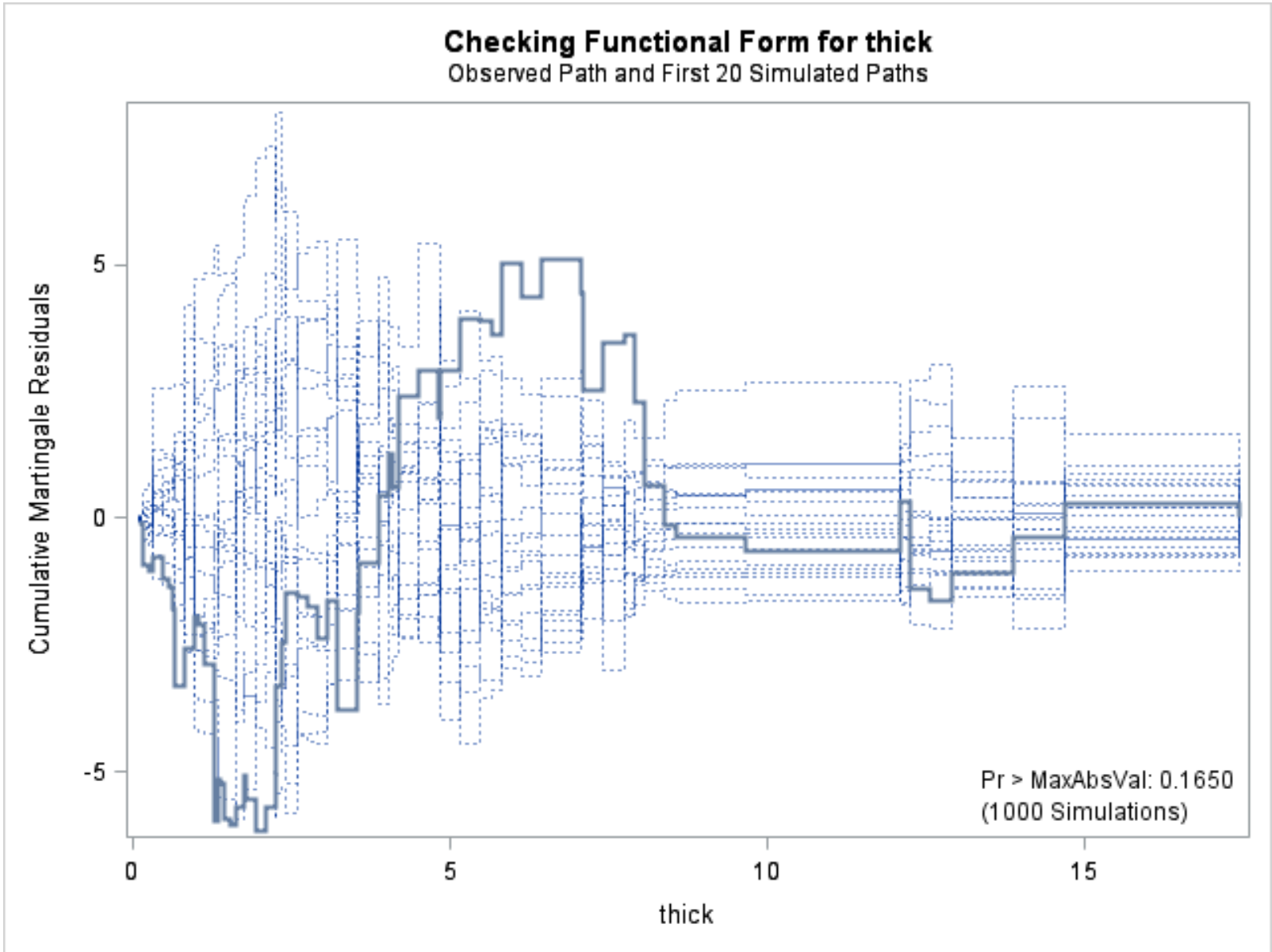
which is 'estimated' as:

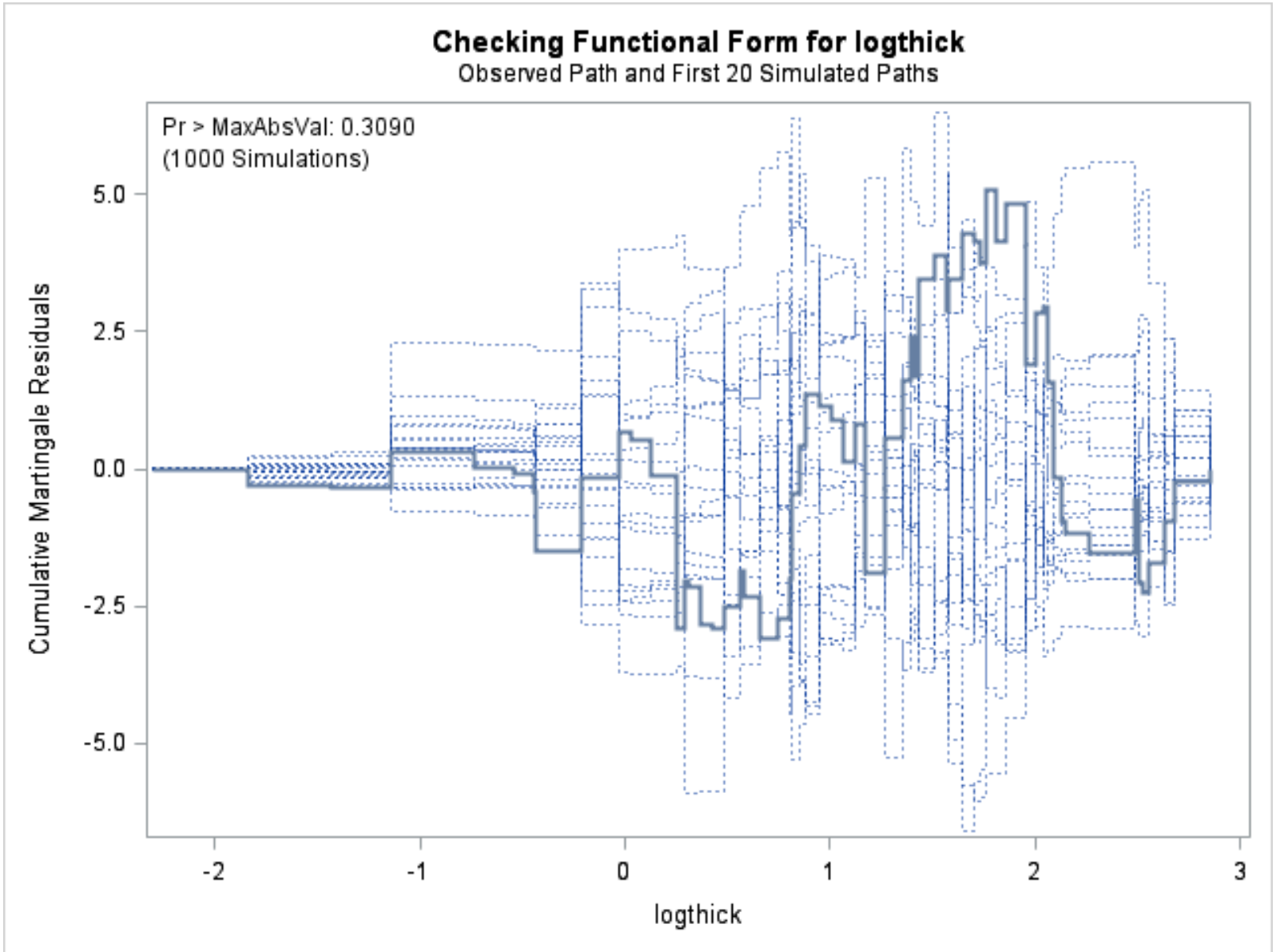
$$\widehat{M}_i(t) = N_i(t) - \int_0^t Y_i(u) \exp(\widehat{\beta}^{\mathsf{T}} X_i) d\widehat{\Lambda}_0(u).$$

Often, the martingale residual is simply defined as $\widehat{M}_i(\infty)$.

Martingale residuals

Martingale residuals may be used directly to check the functional form of a quantitative covariate (e.g., log-linearity): plot $\widehat{M}_i(\infty)$ against X_i and *smooth*.
Plotting *cumulative martingale residuals* against the covariate, it is possible (Lin, Wei and Ying, 1993, *Biometrika*) to get a significance test.
This significance test is based on *re-sampling* from the distribution of the process under the model and evaluating where, in the re-sampled distribution, the observed process is.
We illustrate the method on the melanoma data, death from the disease, tumor thickness.
The technique is available in `PROC PHREG` (`ASSESS` statement) (and in the `R`-package `timereg`).


Checking Functional Form for thick
Observed Path and First 20 Simulated Paths
Cumulative Martingale Residuals
5
0
-5
0
5
10
15
thick
Pr > MaxAbsVal: 0.1650
(1000 Simulations)


Checking Functional Form for logthick
Observed Path and First 20 Simulated Paths
Pr > MaxAbsVal: 0.3090
(1000 Simulations)
Cumulative Martingale Residuals
5.0
2.5
0.0
-2.5
-5.0
-2
-1
0
1
2
3
logthick

Score (Schoenfeld) residuals

The *score* for covariate j is:

$$U_j(\beta, \infty) = \sum_i D_i \Big(X_{ij} - \frac{\sum_{\ell \in R(\widetilde{T}_i)} X_{\ell j} \exp(\beta^{\mathsf{T}} X_\ell)}{\sum_{\ell \in R(\widetilde{T}_i)} \exp(\beta^{\mathsf{T}} X_\ell)} \Big), \quad j = 1, ..., p.$$

The term $U_{ij}(\beta, \infty)$ for subject i (only failures) is the *score-* (or *Schoenfeld-*) *residual*: $U_{ij}(\widehat{\beta}, \infty) = X_{ij} - E_j(\widehat{\beta}, T_i)$.
The *score process* $U_j(\beta, t)$ only adds terms for subjects with $T_i \leq t$.
A *scaled* (or *weighted*) version divides by the estimated variance (say, V_j) of U_j (or by $\sqrt{V_j}$).
A calculation shows the following relation with the martingale residuals:

$$U_j(\widehat{\beta}, t) = \sum_i \int_0^t X_{ij} d\widehat{M}_i(s).$$

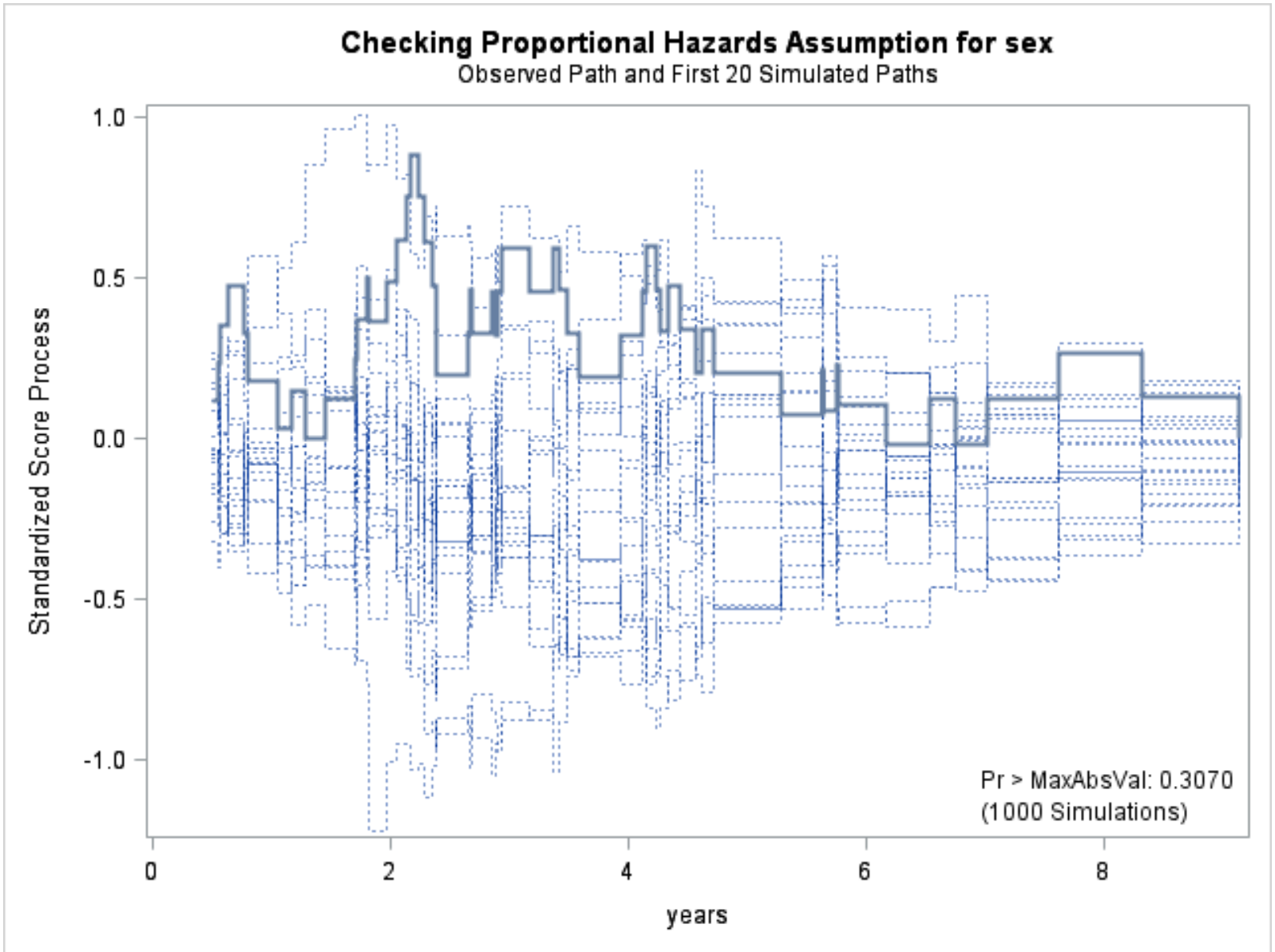
Score (Schoenfeld) residuals

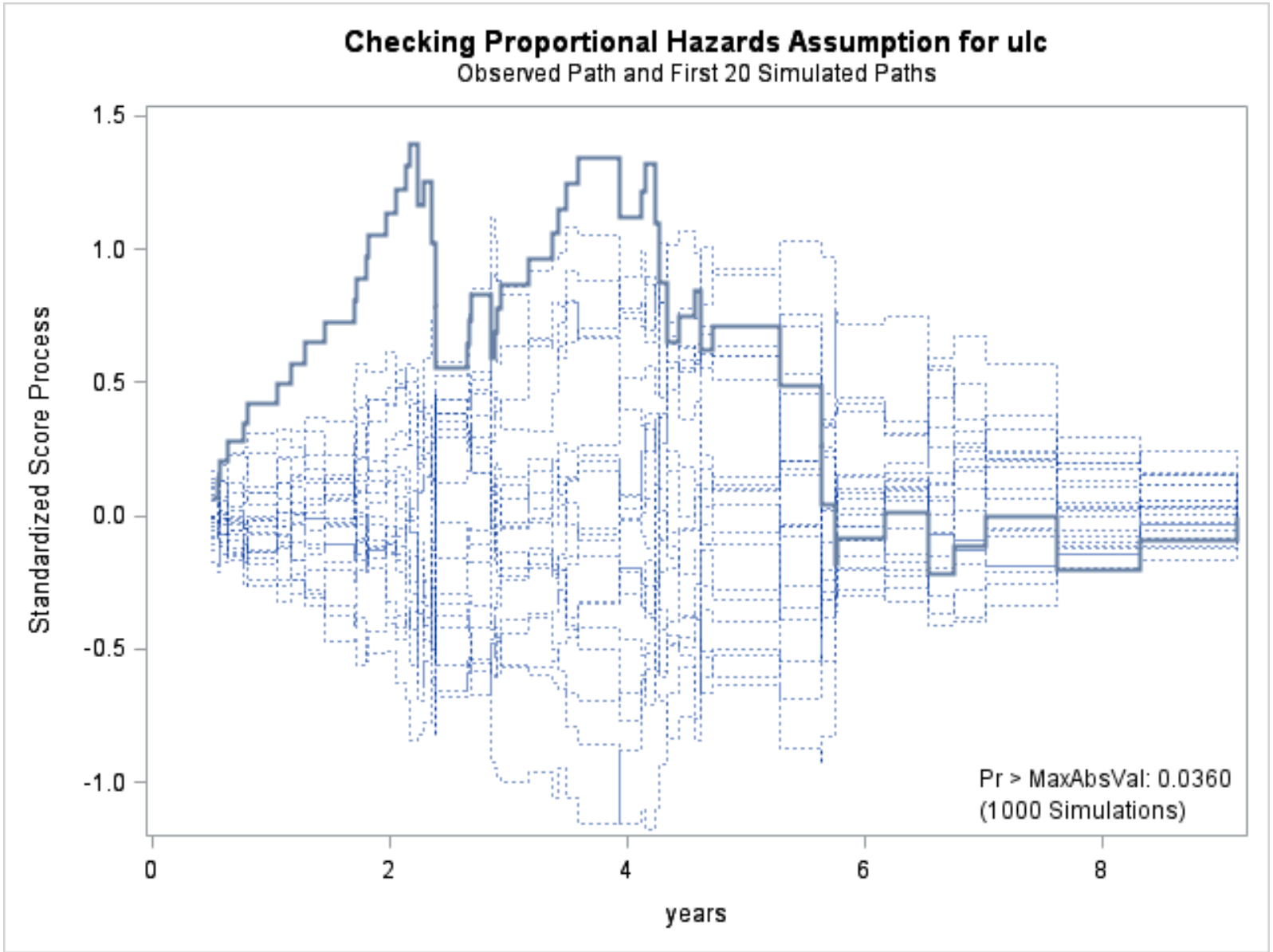
Scaled Schoenfeld residuals may be used directly to check for proportional hazards for a covariate: plot the scaled residual against T_i or against a function $f(T_i)$ and *smooth*. A horizontal curve suggests proportional hazards and a test based on the correlation between $U_{ij}(\widehat{\beta}, \infty)/V_j$ and $f(T_i)$ is available (`ZPH` option in the `PROC PHREG` statement, as well as in `Stata` and in `R`).

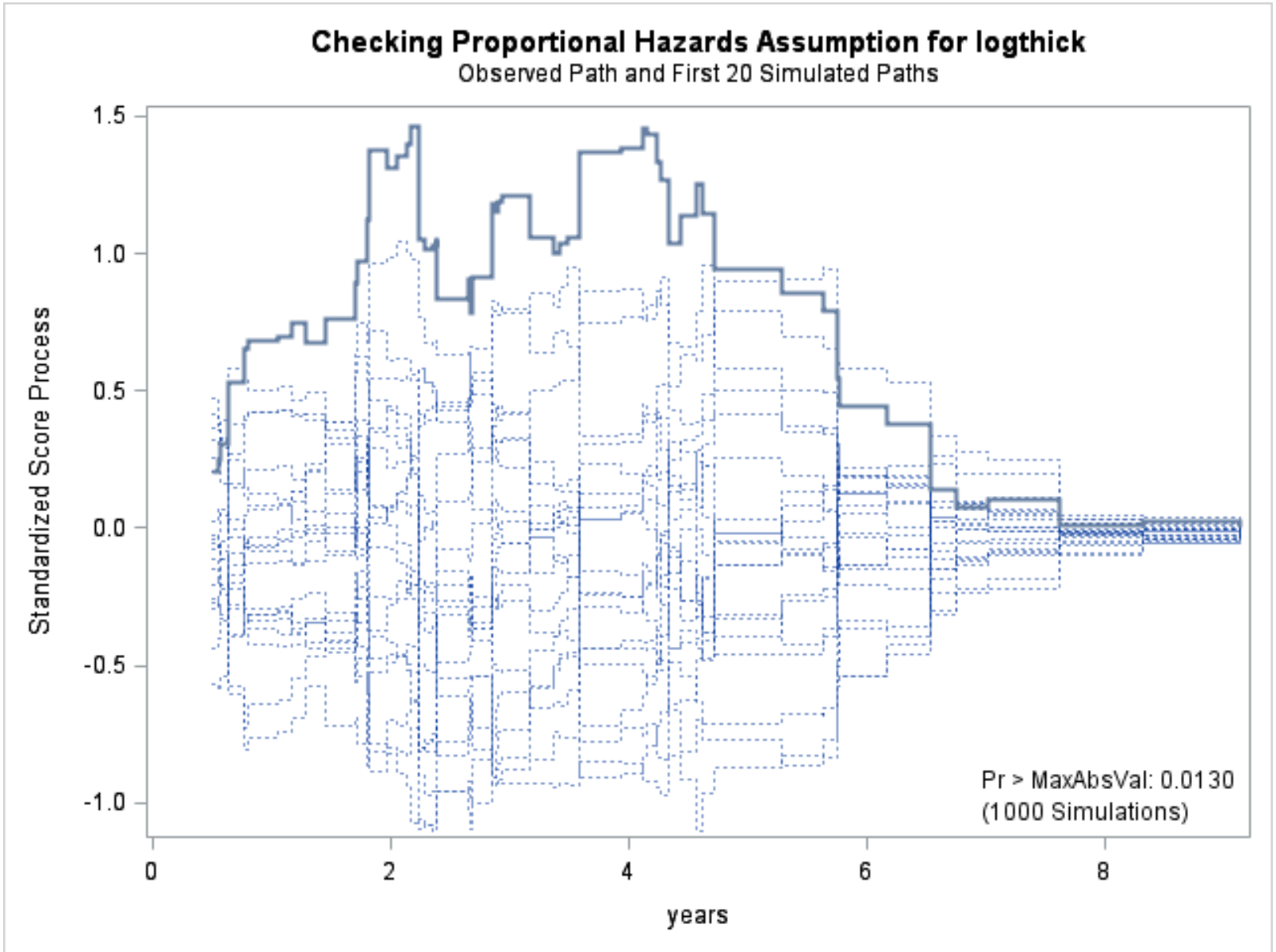
Plotting *cumulative scaled score residuals* (i.e., the scaled score process) against time, it is possible (Lin, Wei and Ying, 1993, *Biometrika*) to get a (better!) significance test.

This significance test is based on *re-sampling* from the distribution of the process under the model and evaluating where, in the re-sampled distribution, the observed process is.

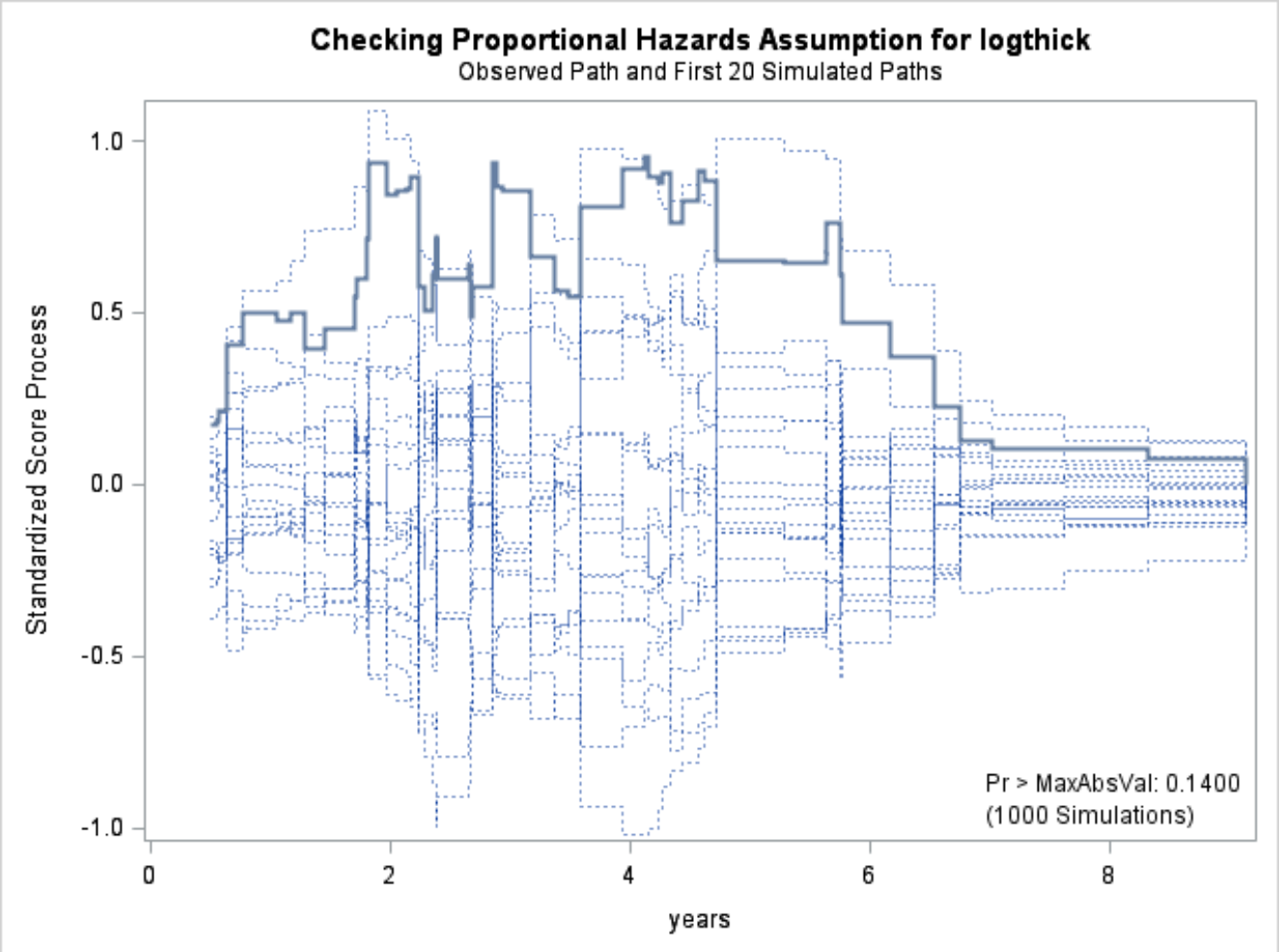
We illustrate the cumulative residual method on the melanoma data, death from the disease.


Checking Proportional Hazards Assumption for sex
Observed Path and First 20 Simulated Paths
Standardized Score Process
1.0
0.5
0.0
-0.5
-1.0
0
2
4
6
8
years
Pr > MaxAbsVal: 0.3070
(1000 Simulations)


Checking Proportional Hazards Assumption for ulc
Observed Path and First 20 Simulated Paths
Standardized Score Process
1.5
1.0
0.5
0.0
-0.5
-1.0
0
2
4
6
8
years
Pr > MaxAbsVal: 0.0360
(1000 Simulations)


Checking Proportional Hazards Assumption for logthick
Observed Path and First 20 Simulated Paths
Standardized Score Process
1.5
1.0
0.5
0.0
-0.5
-1.0
0
2
4
6
8
years
Pr > MaxAbsVal: 0.0130
(1000 Simulations)

Score residuals (stratified model)



Doing it is SAS

```
PROC PHREG DATA = melanoma;
  CLASS ulc sex;
  MODEL days*dc(2 3)=ulc sex logthick / RL;
ASSESS VAR=(logthick) PH / RESAMPLE=1000;
RUN;
```